Spoken term detection (STD) refers to the task of automatically locating the occurrences of a specified query term in a large audio archive. The query term may contain a single word or a sequence of words. It can be given in the form of orthographic representations or query utterance examples. The latter case is known as query-by-example (QbyE) STD. STD technology is useful in various applications, such as multimedia information retrieval, personal entertainment, surveillance and security.

The framework of posteriorgram-based template matching has been shown to be successful for QbyE STD. This framework employs a tokenizer to convert query examples and test utterances into frame-level posteriorgrams, and applies dynamic time warping to match the query posteriorgrams with test posteriorgrams to locate possible occurrences of the query term. It is not trivial to design a reliable tokenizer due to heterogeneous test conditions and the limitation of training resources. This seminar presents a study of using acoustic segment models (ASMs) as the tokenizer. ASMs can be obtained following an unsupervised iterative procedure without any training transcriptions. Details about the ASM training and how to apply ASM for QbyE STD will be introduced. Experimental results show that the ASM tokenizer outperforms a conventional GMM tokenizer and a language-mismatched phoneme recognizer. In addition, the performance is significantly improved by applying unsupervised speaker normalization techniques.